

Natural Language Processing (315329)

MSBTE Diploma · K-Scheme · Model Question Paper — Set B

Time: 3 Hours**Max. Marks: 70****Instructions:**

1. All questions are compulsory.
2. Answer each next main question on a new page.
3. Illustrate your answers with neat sketches wherever necessary.
4. Figures to the right indicate full marks.
5. Assume suitable data, if necessary.
6. Mobile phone, pager and any other electronic communication devices are not permissible in the examination hall.

Q.1 Attempt any FIVE of the following:**10**

- a) State the need of NLP. Name any two areas of linguistics used in NLP.
- b) List any two uses of regular expressions in NLP. Write a regular expression to match a date in dd-mm-yyyy format.
- c) Define tokenization. Give an example of word tokenization.
- d) Define one-hot encoding with an example.
- e) Define polysemy and hyponymy with one example each.
- f) State any two challenges faced while using Transformers.
- g) Define chunking in NLP.

Q.2 Attempt any THREE of the following:**12**

- a) Explain the role of text corpora in NLP. Describe the Brown corpus and the Penn Treebank corpus.
- b) Explain the working of the Porter stemmer. Stem the words: caresses, ponies, connected, hopping, relational, generously.
- c) Draw a semantic network for: “A sparrow is a bird. A penguin is a bird. Birds have wings and can fly, but a penguin cannot fly. Birds are animals.” Explain how it is used.
- d) Find the cosine similarity between the sentences S1: “I like to play cricket and I like to watch cricket” and S2: “I like to watch football” using Bag-of-Words vectors.

Q.3 Attempt any THREE of the following:**12**

- a) Explain the Byte-Pair Encoding algorithm. Perform 4 merges on the corpus: new (6), newer (3), newest (4), few (2).
- b) Explain dependency parsing. Draw the dependency structure of “Rahul bought a new phone yesterday.”

- c) What is lemmatization? How is it different from stemming? Explain with example.
- d) Explain pre-training and fine-tuning in transfer learning. State its benefits.

Q.4 Attempt any THREE of the following:

12

- a) Explain word order typology. Describe SVO, SOV and VSO languages with one example each.
- b) Explain BIO (IOB) tagging used in NER with the example “Narendra Modi visited New York in September”.
- c) Draw the parse tree for the sentence “A boy kicked the red ball.” Write the grammar rules used.
- d) Explain the Bag-of-Words model. Construct the BoW matrix for: D1 “the cat sat on the mat”, D2 “the dog sat on the log”, D3 “the cat chased the dog”.
- e) Write a Python program using NLTK to extract noun phrase, prepositional phrase and verb phrase chunks from “My elder brother bought a red car in Pune”.

Q.5 Attempt any TWO of the following:

12

- a) Explain the n-gram language model. For the corpus “I love NLP”, “I love Python”, “students love NLP”, “I study NLP”, find all bigram probabilities and the probability of the sentence “I love NLP”.
- b) Draw and explain the architecture of the Transformer.
- c) Describe POS tagging. Discuss any four challenges in POS tagging with suitable examples.

Q.6 Attempt any TWO of the following:

12

- a) What are contextual embeddings? Explain how they solve the problem of polysemy, with an example and use cases.
- b) Explain how Hugging Face Transformers are fine-tuned on a custom dataset for sentiment analysis. Include the role of the tokenizer, Datasets and the Trainer API.
- c) Explain the NLP application pipeline for text classification. Explain accuracy, precision, recall and F1-score with a confusion matrix example.

Note: This is a **model question paper prepared by Ananda Technologies** only for practice and exam preparation, based on the pattern of previous MSBTE question papers and the K-scheme syllabus. It is **not an official MSBTE paper**, and it does **not** mean that these questions will come in the exam. Answers follow standard textbook points; marks shown are indicative. Free for students — anandatechnologies.in